

Weekly Report(Nov. 12th, 2018. 刘一璟)

工作

1. china graph 2018参会、报告
2. 将新的降采样方法迁移到vgg16,resnet18上
3. 相似工作的查找、阅读
4. 工作时长: 工作日每日8个小时, 周末共5小时, 共45小时.

工作进度

项目	进度	截止时间
投稿	2.代码部分已经完成, 接下来希望在coco等数据集上进行实验3.除空间变换网络外, 17年ICML的最佳论文网络可解释的工作也与新的方法有一点关联. 直接从降采样方式入手的相关工作并没有找到, 上面两个工作与新的方法相关性不大. 需要继续寻找与新方法相关的工作.	待定

论文阅读

Spatial Transformer Network

引入了一个新的可学模块, 空间变换网络, 它显式地允许在网络中对数据进行空间变换操作。这个可微的模块可以插入到现有的卷积架构中, 使神经网络能够主动地在空间上转换特征映射。

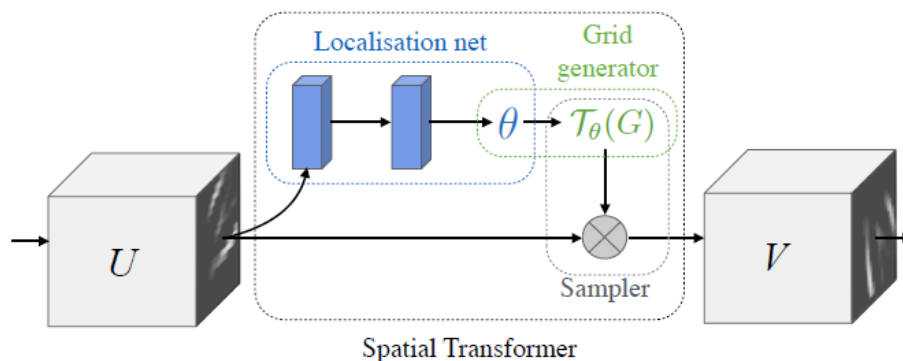


Figure 2: The architecture of a spatial transformer module. The input feature map U is passed to a localisation network which regresses the transformation parameters θ . The regular spatial grid G over V is transformed to the sampling grid $\mathcal{T}_\theta(G)$, which is applied to U as described in Sect. 3.3, producing the warped output feature map V . The combination of the localisation network and sampling mechanism defines a spatial transformer.

Understanding Black-box Predictions via Influence Functions

这篇论文提出了找出图片中哪些点对最终分类结果有主要影响的方法，属于深度学习可解释性的工作。与我们提出的新方法相似之处在于，新方法同样需要对图片不同区域进行类似的重要程度评估，并保留更为重要的区域。

主要思想是：在测试点周围拟合一个简单模型，然后扰动测试集，看预估值如何变化。

Influence functions

- Goal: Measure change in $L(z_{\text{test}}, \hat{\theta}_{\epsilon, z_{\text{train}}})$ as we increase ϵ .
- $\hat{\theta}_{\epsilon, z_{\text{train}}} \stackrel{\text{def}}{=} \arg \min_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n L(z_i, \theta) + \epsilon L(z_{\text{train}}, \theta)$.
- Under smoothness assumptions,

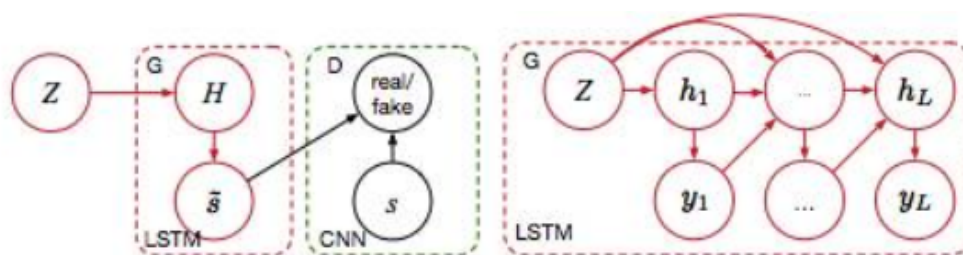
$$\begin{aligned} \mathcal{I}_{\text{up, loss}}(z_{\text{train}}, z_{\text{test}}) &\stackrel{\text{def}}{=} \left. \frac{dL(z_{\text{test}}, \hat{\theta}_{\epsilon, z_{\text{train}}})}{d\epsilon} \right|_{\epsilon=0} \\ &= -\nabla_{\theta} L(z_{\text{test}}, \hat{\theta})^{\top} H_{\hat{\theta}}^{-1} \nabla_{\theta} L(z_{\text{train}}, \hat{\theta}), \end{aligned}$$

- where $H_{\hat{\theta}} \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^n \nabla_{\theta}^2 L(z_i, \hat{\theta})$.

Generating Text via Adversarial Training

尝试将 GAN 理论应用到了文本生成任务上。文中的方法比较简单，具体可以总结为：

以递归神经网络（LSTM）作为GAN的生成器（generator）。其中，用光滑近似（smooth approximation）的思路来逼近 LSTM 的输出。结构图如下：



本文生成模型 (LSTM) decode阶段有exposure bias问题，即在训练过程中逐渐用预测输出替代实际输出作为下一个词的输入